

The demand must be filed directly with the competent International Preliminary Examining Authority or, if two or more Authorities are competent, with the one chosen by the applicant. The full name or two-letter code of that Authority may be indicated by the applicant on the line below:

IPEA/ EPO

PCT

CHAPTER II

DEMAND

under Article 31 of the Patent Cooperation Treaty:

The undersigned requests that the international application specified below be the subject of international preliminary examination according to the Patent Cooperation Treaty.

For International Preliminary Examining Authority use only

Identification of IPEA		Date of receipt of DEMAND	
Box No. I IDENTIFICATION OF THE INTERNATIONAL APPLICATION		Applicant's or agent's file reference 800293WO	
International application No. PCT/FI2004/000331	International filing date (day/month/year) 01 June 2004 (01.06.2004)	(Earliest) Priority date (day/month/year) 06 June 2003 (06.06.2003)	
Title of invention Processing data records for finding counterparts in a reference data set			
Box No. II APPLICANT(S)			
Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.) TIETOENATOR OYJ Kutojantie 10 FI-02630 Espoo Finland		Telephone No. +358 9 862 6000	
		Facsimile No. +358-9 8625 2597	
		Teleprinter No.	
		Applicant's registration No. with the Office	
State (that is, country) of nationality: FI		State (that is, country) of residence: FI	
Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.) SURAKKA, Timo Länsituulentie 8 B 29 FI-02100 Espoo Finland			
State (that is, country) of nationality: FI		State (that is, country) of residence: FI	
Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.) PULKKINEN, Pentti Meiramikuja 3 B 6 FI-02940 Espoo Finland			
State (that is, country) of nationality: FI		State (that is, country) of residence: FI	
<input type="checkbox"/> Further applicants are indicated on a continuation sheet.			

Box No. III AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCEThe following person is ☒ agent ☐ common representativeand ☒ has been appointed earlier and represents the applicant(s) also for international preliminary examination.☐ is hereby appointed and any earlier appointment of (an) agent(s)/common representative is hereby revoked.☐ is hereby appointed, specifically for the procedure before the International Preliminary Examining Authority, in addition to the agent(s)/common representative appointed earlier.Name and address: *(Family name followed by given name; for a legal entity, full official designation.
The address must include postal code and name of country.)*KUISMA, Sirpa
Page White & Farrer
Runeberginkatu 5, 10th Floor
FIN-00100 Helsinki
Finland

Telephone No.

+358 9 343 6510

Facsimile No.

+358 9 343 65111

Teleprinter No.

Agent's registration No. with the Office

☐ Address for correspondence: Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.**Box No. IV BASIS FOR INTERNATIONAL PRELIMINARY EXAMINATION****Statement concerning amendments:***

1. The applicant wishes the international preliminary examination to start on the basis of:

☐ the international application as originally filedthe description ☒ as originally filed☐ as amended under Article 34the claims ☐ as originally filed☐ as amended under Article 19 (together with any accompanying statement)☒ as amended under Article 34the drawings ☒ as originally filed☐ as amended under Article 342. ☐ The applicant wishes any amendment to the claims under Article 19 to be considered as reversed.3. ☐ The applicant wishes the start of the international preliminary examination to be postponed until the expiration of the applicable time limit under Rule 69.1(d).4. ☐ The applicant expressly wishes the international preliminary examination to start earlier than at the expiration of the applicable time limit under Rule 54bis.1(a).

* Where no check-box is marked, international preliminary examination will start on the basis of the international application as originally filed or, where a copy of amendments to the claims under Article 19 and/or amendments of the international application under Article 34 are received by the International Preliminary Examining Authority before it has begun to draw up a written opinion or the international preliminary examination report, as so amended.

Language for the purposes of international preliminary examination: English

☒ which is the language in which the international application was filed.☐ which is the language of a translation furnished for the purposes of international search.☒ which is the language of publication of the international application.☐ which is the language of the translation (to be) furnished for the purposes of international preliminary examination.**Box No. V ELECTION OF STATES**

The filing of this demand constitutes the election of all Contracting States which are designated and are bound by Chapter II of the PCT.

Box No. VI CHECK LIST

The demand is accompanied by the following elements, in the language referred to in Box No. IV, for the purposes of international preliminary examination:

- | | | |
|--|---|----------|
| 1. translation of international application | : | sheets |
| 2. amendments under Article 34 | : | 5 sheets |
| 3. copy (or, where required, translation) of amendments under Article 19 | : | sheets |
| 4. copy (or, where required, translation) of statement under Article 19 | : | sheets |
| 5. letter | : | 3 sheets |
| 6. other (<i>specify</i>) courtesy copy of amendments | : | 5 sheets |

For International Preliminary Examining Authority use only

received	not received
----------	--------------

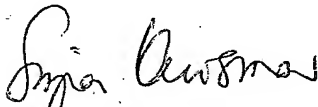
<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>

The demand is also accompanied by the item(s) marked below:

- | | |
|--|--|
| 1. <input checked="" type="checkbox"/> fee calculation sheet | 5. <input type="checkbox"/> statement explaining lack of signature |
| 2. <input type="checkbox"/> original separate power of attorney | 6. <input type="checkbox"/> sequence listing in computer readable form |
| 3. <input type="checkbox"/> original general power of attorney | 7. <input type="checkbox"/> tables in computer readable form related to a sequence listing |
| 4. <input type="checkbox"/> copy of general power of attorney; reference number, if any: | 8. <input type="checkbox"/> other (<i>specify</i>): EPO Form 1037 |

Box No. VII SIGNATURE OF APPLICANT, AGENT OR COMMON REPRESENTATIVE

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the demand).



Sirpa Kuisma, Professional representative
Page White & Farrer

For International Preliminary Examining Authority use only

1. Date of actual receipt of DEMAND:

2. Adjusted date of receipt of demand due to CORRECTIONS under Rule 60.1(b):

3. ☐ The date of receipt of the demand is AFTER the expiration of 19 months from the priority date and item 4 or 5, below, does not apply.
☐ The applicant has been informed accordingly.
4. ☐ The date of receipt of the demand is WITHIN the time limit of 19 months from the priority date as extended by virtue of Rule 80.5.
5. ☐ Although the date of receipt of the demand is after the expiration of 19 months from the priority date, the delay in arrival is EXCUSED pursuant to Rule 82.

6. ☐ The date of receipt of the demand is AFTER the expiration of the time limit under Rule 54bis.1(a) and item 7 or 8, below, does not apply.
7. ☐ The date of receipt of the demand is WITHIN the time limit under Rule 54bis.1(a) as extended by virtue of Rule 80.5.
8. ☐ Although the date of receipt of the demand is after the expiration of the time limit under Rule 54bis.1(a), the delay in arrival is EXCUSED pursuant to Rule 82.

For International Bureau use only

Demand received from IPEA on:

10/559386

IAP16 Rec'd PCT/PTO 06 DEC 2004



5 April 2005

**Page White
and Farrer**

Patent and
Trade Mark Attorneys

Finland
Runeberginkatu 5
10th Floor
FIN-00100 Helsinki
Tel +358 9 343 6510
Fax +358 9 343 65111
helsinki@pagewhite.com

UK
54 Doughty Street
London
WC1N 2LS
Tel +44(0)20 7831 7929
Fax +44(0)20 7831 8040
london@pagewhite.com

www.pagewhite.com

European Patent Office
Erhardtstrasse 27
D-80331 München
Germany

Dear Sirs

Re: International Patent Application No. PCT/FI2004/000331
Tietoenator Oyj
Our ref: 800293WO

In connection with filing of the Demand, we hereby submit amended claims and respectfully present the following in response to the Written Opinion of the International Searching Authority issued on 13 October 2004.

The independent claims 1, 21, 26 and 28 are hereby amended to define that a synonym acceptance criterion takes into account the number of identical characters; support for the amendment can be found, for example, on page 1, lines 5-7 and page 20, lines 17-20. Claim 1 has also been amended to state that the counterparts are searched for after determining synonym candidates. The enclosed dependent claims remain unchanged.

The present invention addresses processing of data records containing data fields. In other words, the present invention handles structured data. The present invention aims to increase the accuracy of finding correct counterparts in a reference data set for processed data records. The search for the counterpart in the reference set typically takes into account a number of data fields in the data record.

The meaning of the identifier, which a data field represents, is irrelevant for the present invention. The claimed invention is concerned about the value of the data field, for example, about the character string in the data field. The amended independent claims make this clear by stating that the synonym acceptance criterion takes into account the number of identical characters in the synonym candidate and in the value of the data field.

In the claimed invention, there is a set of predetermined identifier values and a predefined synonym acceptance criterion. At least one synonym candidate is determined for a data field value from the set of predetermined identifier values and, if the data field value and the synonym candidate fulfill the synonym acceptance criterion, the synonym candidate and the data field value are associated as synonyms.

Directors:
P D Jenkins
Mrs V R Driver
J N Daniels
Ms K C Style
Ms N Shackleton
P R Slingsby
C M Hill
J P Ruuskanen
J P Cornish
D J Williams

Associate Directors:
M N Evans
Ms J H Evenson
Ms C A Wolfe

Ms A Campbell
Ms S Kuisma
Ms V Tognetty
M Levin
R Dempster

Consultant:
D J Richards

FI Register No. Y-1749026-1

This means that information about the synonyms is updated, if the synonym acceptance criterion is fulfilled, based on the current value of the data field. The update of the information about synonyms occurs before the search for the counterpart in the reference data set. As the synonym information has been updated based on the value of a data field in the current data record, it is more likely that a counterpart is found for the data record than when using static synonym information.

The updating of the information about synonyms may be automated, as there exists the set of predefined identifier values and a predetermined synonym acceptance criterion is defined.

D1 (Sa Lin et al) discloses methods for overcoming problems relating to the use of databases when the user's terminology does not match the terminology used in the database. The solution is to describe the terms and the relationships between the terms (that is, to construct an ontology model) and to expand user search terms using the ontology model.

Ontology refers to the meaning of terms/words. The ontology model in D1 is typically built in advance. In D1 search terms are thus expanded before a search based on the meaning of the search terms using a (typically) static ontology model.

The claimed invention is therefore different from the disclosure of D1 and thus novel in view of D1.

Furthermore, the aim of the method discussed in D1 is different from the aim of the invention. In D1, the aim is to expand the search terms for finding larger amounts of relevant information from the available information. A person skilled in the art of processing structured information, when trying to increase the accuracy of finding counterparts for data records, would not consult publications relating to handling terms based on the meaning of the terms. As mentioned above, the meaning of a term (identifier) is irrelevant for the present invention.

In addition, should the skilled person have a look at D1, he would notice that the ontology model in D1 is static or it is updated based on user input. In D1 it is expressly said that attempts to completely automate update of the ontology model have not been very promising (page A33, left column, above Figure 1).

Thus D1 would not lead a skilled person into defining a synonym acceptance criterion and updating information about synonyms before searching counterparts for a data record. We thus find the claimed invention inventive in view of D1.

Regarding the other documents cited in the International Search Report, we would like to mention the following.

WO0141002 relates to use of distributed databases. The user is allowed to make unstructured queries, and the query terms are generalized and/or expanded to return as many relevant words as possible to the user. Similarly as in D1, also in this publication the search terms are processed based on their meaning.

The publication by Rodriguez and Varas (XP-002297455) discusses ontologies and database schemas, similarly as D1. In this paper there is no hint to updating a database schema.

The publication by Luján-Mora and Palomar addresses integration of data from different sources. The proposed solution is based on clustering: various variants of a term are replaced by the most frequently occurring variant.

This publication addresses handling of structured data, similarly as the present invention. The solution to cope with various variants of a term is based on the assumption that the most frequently occurring variant is correct. There is no explicit knowledge about the correct variant, in contrast to the present invention where there is a set of predefined identifier values.

Based on the above, we find the claimed invention new and inventive over the cited prior art. A reconsideration of the statement regarding novelty and inventive step in the Written Opinion of the International Searching Authority is therefore respectfully requested.

Should the Examiner, despite the enclosed amended claims and the arguments presented above, consider issuing a negative International Preliminary Examination report, the applicant expects to receive a further Written Opinion pursuant to Rule 66.2 PCT.

Please acknowledge receipt of this letter by returning the top copy of the EPO Form 1037 enclosed with this letter.

Yours faithfully



Sirpa Kuisma
Professional Representative

Encl. - replacement pages 23-27
- courtesy copy of pages 23-27 showing amendments of the independent claims

Claims

1. A method of processing a data record for finding a counterpart in a reference data set, the method comprising the steps of:

5 determining in the data record a value of a data field, the data field representing an identifier,

 determining from a set of predetermined identifier values at least one synonym candidate for the value of the data field,

10 determining if a synonym candidate and the value of the data field fulfill a predetermined synonym acceptance criterion taking into account the number of identical characters in the synonym candidate and in the value of the data field, and if the predetermined synonym acceptance criterion is fulfilled, associating the value of the data field and the synonym candidate as synonyms, and

15 searching for a counterpart for the data record by comparing to entries of the reference data set the value of the data field and/or a synonym associated with the value of the data field after determining whether the predetermined synonym acceptance criterion is fulfilled.

20 2. A method as defined in claim 1, wherein the at least one synonym candidate is determined using a candidate selection criterion depending at least on the value of the data field and on a synonym candidate.

25 3. A method as defined in claim 2, wherein the candidate selection criterion takes into account how similar a synonym candidate and the value of the data field sound.

4. A method as defined in claim 2, wherein the candidate selection criterion specifies that at least a predetermined part of the value of the data field is identical to a predetermined part of a synonym candidate.

30 5. A method as defined in any one of claims 2 to 4, wherein the candidate selection criterion takes into account also a further data field of the data record, said further data field representing a second identifier.

6. A method as defined in any preceding claim, wherein at least one quality parameter is evaluated for a synonym candidate, the synonym acceptance criterion taking into account the at least one quality parameter.

5 7. A method as defined in claim 6, wherein at least one quality parameter takes into account at least one of the following quantities:

a number of changes required for converting the value of the data field to be identical to a synonym candidate; a proportion of identical characters in the value of the data field and in a synonym candidate; and a difference between the length of the value of
10 the data field and the length of a synonym candidate.

8. A method as defined in claim 7, wherein the number of changes required for converting the value of the data field to be identical to a synonym candidate is calculated using the Levenshtein distance.

15

9. A method as defined in claim 7, wherein the proportion of identical characters takes into account the order of the characters.

10. A method as defined in any one of claims 6 to 9, wherein a first quality parameter
20 is evaluated for each synonym candidate and at least a second quality parameter is evaluated at least for the synonym candidate(s) having the best first quality parameter.

11. A method as defined in any one of claims 6 to 10, wherein the synonym acceptance criterion requires that there is only one synonym candidate having the best
25 at least one quality parameter.

12. A method as defined in any one of claims 6 to 11, wherein at least two quality parameters are evaluated for each synonym candidate and the synonym candidate acceptance criterion specifies a threshold for one of the at least two quality
30 parameters, the threshold being dependent on a further one of the at least two quality parameters.

13. A method as defined in any preceding claim, wherein the search for the counterpart involves comparison of the value of the data field to a synonym set

relating to the identifier, members of said synonym set referring to respective predetermined identifier values, and when the predetermined synonym acceptance criterion is fulfilled, the value of the data field is added to the synonym set as a member referring to the synonym associated with the value of the data field before the search for the counterpart.

14. A method as defined in any preceding claim, wherein determining the at least one synonym candidate is discarded, if a predetermined discard criterion is fulfilled.

15. A method as defined in claim 14, wherein the predetermined discard criterion specifies that the value of the data field is identical to one of the predetermined identifier values.

16. A method as defined in claim 14, wherein the search for the counterpart involves the synonym set and the predetermined discard criterion specifies that the value of the data field is at least one of the following: one of the predetermined identifier values, and a member of the synonym set.

17. A method as defined in any one of claims 14 to 16, wherein the predetermined discard criterion takes into account a value of a second data field in the data record.

18. A method as defined in any preceding claim, wherein information indicating the at least one synonym associated with the value of the data field is added to the data record.

19. A method as defined in claim 18, wherein a copy of the data record is made for each synonym associated with the value of the data field.

20. A method as defined in any preceding claim, wherein the identifier relates to a name of one of the following: a geographical entity, a person and an organisation.

21. A method of processing a synonym set for searching counterparts in a reference data set for data records, a data record containing a data field representing an identifier, members of the synonym set being first identifier values and referring to

respective second identifier values, the second identifier values being predetermined identifier values, and said searching for a counterpart involving comparison of a value of the data field to the synonym set, the method comprising the steps of determining among the predetermined identifier values at least one synonym candidate relating to
 5 the value of the data field in the data record, and, if the value of the data field and a synonym candidate fulfill a predetermined synonym acceptance criterion taking into account the number of identical characters in the synonym candidate and in the value of the data field, adding before searching a counterpart for a data record the value of the data field to the synonym set as a member referring to the synonym candidate.

10

22. A method as defined in claim 21, wherein the synonym set is empty before adding the value of the data field to the synonym set.

15

23. A method as defined in claim 21, wherein the synonym set contains at least one member before adding the value of the data field to the synonym set.

24. A computer program comprising program instructions for causing a computer to perform the method of any one of claims 1 to 23.

20

25. A computer program as defined in claim 24, embodied on a computer-readable record medium.

26. A data processing system for processing data records for finding counterparts in a reference data set, the system comprising:

25

- means for receiving data records,
- means for storing the reference data set,
- means for storing predetermined identifier values for an identifier,
- means for determining in the data records values of a data field, the data field representing the identifier,

30

- means for associating values of the data field and respective predetermined identifier values as synonyms, said means configured to determine from the predetermined identifier values at least one synonym candidate for a value of the data field, to determine if a synonym candidate and the value of the data field fulfill a predetermined synonym acceptance criterion taking into account the

number of identical characters in the synonym candidate and in the value of the data field, and if the predetermined synonym acceptance criterion is fulfilled, to associate the value of the data field and the synonym candidate as synonyms, and

- means for searching counterparts in the reference data set for the data records, said searching involving comparing to entries of the reference data set values of data fields and/or synonyms associated with the values of the data fields.

27. A data processing system as defined in claim 26, further comprising

- means for storing a synonym set, members of said synonym set referring to respective predetermined identifier values,

wherein the means for associating values of the data field and respective predetermined identifier values as synonyms are configured to add to the synonym set a member referring to the synonym associated with the value of the data field before activation of the means for searching counterparts.

28. A data processing system for processing a synonym set for searching counterparts in a reference data set for data records, a data record comprising a data field representing an identifier, members of the synonym set being first identifier values and referring to respective second identifier values, said second identifier values being predetermined identifier values, and said searching involving comparing a value of the data field to the synonym set, the system comprising:

- means for storing the synonym set,
- means for storing predetermined identifier values for the identifier,
- means for receiving data records,
- means for determining in the data records values of the data field, and
- means for adding to the synonym set a value of the data field and respective predetermined identifier values associated as synonyms before searching counterparts in the reference data set, said means configured to determine from the predetermined identifier values at least one synonym candidate for a value of the data field, to determine if a synonym candidate and the value of the data field fulfill a predetermined synonym acceptance criterion taking into account the number of identical characters in the synonym candidate and in the value of the data field, and if the predetermined synonym acceptance criterion is fulfilled, to associate the value of the data field and the synonym candidate as synonyms.